

Evaluating Data Observability Platforms

How to assess your data observability needs
and frameworks



INTRODUCTION

What to consider

If you're considering a data observability solution, how do you know what to look for? We built this step-by-step Data Observability Request For Information (RFI) after running hundreds of Proofs of Concepts (POCs). Within, you'll find the end result of thousands of hours and dozens of experts and customers, distilled into the best way to run a thorough data observability evaluation.

The desire to be data-driven is nothing new. Organizations have been striving for “data-drivenness” for a couple of decades. New tools have helped, but many organizations still sit on a big stack of data they either can't use, or don't know how to leverage.

If you asked teams why they're looking to invest in data observability, they might say:

“I don't know what this data is or how I'm supposed to use it.”

“We have way too much data and I don't even know where to start.”

“I'm not confident the data is reliable.”

Many companies have tried to solve these issues over the past decade. But “I'm not confident the data is reliable” continues to plague even the most sophisticated DataOps teams. There's just too much data, too many potential points of failure, and too few skilled humans to test all of it.

Is data observability right for you?

Probably, if you've recently asked yourself or your team:

“How many more failed data merge attempts can I debug before I do something I'll regret?”

“How do I find data issues before my executives or customers do?”

“How can I make sure my replication job completes on time and all the data shows up looking like I meant it to?”

In general, when your teams start investing serious time and attention into recurring data quality conundrums, it's time to make a change. Luckily, there is a better way forward. Data observability tools provide the automation, consistency, and peace of mind you need with respect to your data pipelines.

3 steps to evaluating data observability solutions

1 Assemble your team

An evaluation team that can gather requirements and synthesize them before you begin looking at vendors. That team might be made up of:

- **Practitioners:** One or two representatives from each team who might interact with data observability. This could include data engineers, data scientists, ML engineers, or analysts.
- **Evaluation leader:** A project leader who's close to both the technical work and the business requirements—such as a director of data engineering or a seasoned engineer.
- **Executive sponsor:** A CTO or CDO, whose buy-in will go a long way in aligning with your organization's strategic goals and securing budget.
- **Purchasing:** Post-evaluation, procurement and purchasing can take much longer than many people think. Identify leaders for legal reviews, procurement steps, etc.

Also consider a cross section of teams who will use the data observability platform:

- **Data engineering / data platform:** The core users of your data observability solution, who should be core decision makers in the selection process.
- **Analytics / data science:** Subject matter experts and heavy users of data observability, once pipeline freshness / volume monitoring is set up.
- **Machine learning:** These teams often adapt data observability to protect the pipelines feeding the training and serving data used to drive models that reach production.

EXAMPLE TEAM LIST

Executive Sponsor



Michelle C.
Head of Data

POC Leader



John H.
Data Engineering
Manager

Technical Evaluators



Alan G.
Analytics team 1



Ellie S.
Analytics team 2



Ian M.
ML-Eng team

Purchasing



L. Dodgson
Legal



J. Arnold
Procurement

2 Gather your requirements

Your requirements are the laundry list of wants and needs for your business. Tie each list item to a specific business challenge or initiative. Look at these data observability categories for inspiration:

Deployment: Are you a nimble startup in need of a fully-managed SaaS solution? A large healthcare provider looking for the flexibility of a single-tenant hosted option? Consider how the solution will connect to your existing stack and what performance impacts it may have on your operation.

Monitoring: Monitoring capabilities are the foundation of data observability. Use the requirements you outlined to determine which types of monitoring you'll need and how much effort will be required to deploy and maintain them. Which monitors are available out of the box and which will need to be custom?

Resolution: Knowing about a problem is great, knowing how to resolve it is even better. How do you track and manage issues, and is root cause analysis and debugging included?

Security: Engage your IT security team early to understand your organization's software security requirements: for example SOC2 certification, compliance with relevant industry regulations, and whether the vendor exports or stores any of your data.

Integrations: Evaluate the current tools in your data stack and consider how a data observability solution will interact with—and add value to—each. These may include analytics tools like Tableau, data catalogs like Alation, and orchestration and transformation tools like Fivetran, dbt and Airflow.

Alerts and notification: You don't want to be overloaded with false positives or non-issues. Your solution should provide an optimized alerting system where you can easily fine-tune notifications.



3 STEPS TO EVALUATING DATA OBSERVABILITY SOLUTIONS

Reporting: Reporting provides insight into performance over time. Do reports include which pipelines are currently in or out of coverage, how many issues have been detected / acknowledged / resolved, and if issues are trending over time?

Anomaly detection: Anomaly detection should help you identify issues you didn't know existed, a capability very difficult to build yourself. Look at what types of anomaly detection are available and the sophistication of the ML models used.

Configuration and management: Configuring and management is critical to your success as data observability usage and requirements grow. Look for simple, scalable, Terraform-like “as code” options.



3 Selecting vendors to evaluate

Your solution's requirements are ultimately driven by the business requirements you just defined. In general, you want to look for vendors that provide a unified platform, where all stakeholders can communicate in a centralized location. Automation is another non-negotiable: intelligent alerts and incident prioritization should be a main feature. Look for end-to-end adaptability that can grow and scale as your business changes. Lastly, look for robust security and compliance components. Additionally, here's a bit more about vendor preferences, based on the makeup of your team:

If your team is primarily data engineers, they'll want:

- To know that data in the warehouse is fresh and complete
- The ability to query the metadata and expose broad operational health metrics across the data warehouse
- Little to no setup of management on your part

If you have data science and analytics teams, they'll want:

- To ensure data adheres to business-logic quality rules before it's fed into an analytics dashboard or ML application
- Deep business-logic driven metrics such as variance or outlier detection



Want to run a great POC?

Skip the work and download our free Data Observability RFI (Request for Information). It contains a comprehensive checklist of features, and makes it easy to add additional criteria to score vendors on, so your team can make the best decision.

- ✓ Pre-built data observability features list
- ✓ Grouped capabilities to help you focus on what matters most to you
- ✓ Compare multiple vendors during your evaluation

Download now 



Reliable data starts here

Find and fix problems anywhere in your data pipeline, before users find out.